



CHICAGO JOURNALS



History  
of  
Science  
Society

---

Telepathy: Origins of Randomization in Experimental Design

Author(s): Ian Hacking

Source: *Isis*, Vol. 79, No. 3, A Special Issue on Artifact and Experiment (Sep., 1988), pp. 427-451

Published by: [The University of Chicago Press](#) on behalf of [The History of Science Society](#)

Stable URL: <http://www.jstor.org/stable/234674>

Accessed: 22/05/2014 14:22

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*The University of Chicago Press* and *The History of Science Society* are collaborating with JSTOR to digitize, preserve and extend access to *Isis*.

<http://www.jstor.org>

# Telepathy: Origins of Randomization in Experimental Design

*By Ian Hacking\**

## I. RANDOMIZATION AT PRESENT

PSYCHIC RESEARCH may seem an implausible place to study the emergence of a new kind of experimental methodology. Yet it is there that we find the first faltering use, by many investigators, of a technique that is now standard in many sciences and mandatory in much sociology and biology. Indeed randomization is so commonplace that anyone untroubled by the fundamental principles of statistics must suppose that the practice is quite uncontroversial. That is not so. Before I proceed to my story, a brief account of experimental design in the twentieth century is in order. Modern advocates of an almost universal use of randomization in experimental design may find my account of its origins merely curious. But opponents of randomization as a universal practice may well say that it should be restricted to precisely the kinds of work where it started: work marked chiefly by complete ignorance.

There is nothing demeaning about ignorance. There is an important class of empirical experiments in which we want to know whether a treatment of a certain kind makes a difference. "Empirical experiment" is not, as it sounds, a tautology. I refer to experiments in which we have only a very modest understanding of what is going on, as when we administer a drug to a patient, a fertilizer to a field, a poison to an insect. Of course we have some reason to suspect that the treatment may make a difference, although often, if one takes the time to wade through summaries and abstracts, it is only a reason of the form "Chemicals like this seem sometimes to have some effect on problems like that; let's see what happens."

It is now taken for granted, at least in the popular mind, that when possible experimental procedure should proceed as follows. First a number of "patients"—I shall take the case of human subjects—are "randomly" divided into two lots, those that will receive treatment and those that will not. However, the subjects

\* Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto, Canada M5S 1K7.

This paper was written at the Institute for Advanced Study, Princeton, where I was supported by an Izaak Walton Killam Research Fellowship (Canada Council) supplemented by funds from the Henry Luce Foundation (IAS). Davis Baird, Lorraine Daston, Persi Diaconis, Stephen Stigler, and Zeno Swijtink read an earlier draft and corrected errors and provided new information. The editorial staff of *Isis* provided invaluable assistance.

are “blind.” No one of them knows whether he or she is singled out for treatment. Those who are not treated are given a placebo perceptually indistinguishable from the drug. Moreover, the experiment is double-blind, with the experimenter not knowing which patient is receiving treatment and observing only how well the members of the entire population fare. Only after assessment is the experimental blindfold removed for final analysis.

That is the most favored case. Many treatments are not a matter of giving tablets. In comparing two kinds of surgery for breast cancer no one can be blind, and we have ethical problems about assigning one of the two modes of surgery to a patient “at random.” At the very least, she should have some choice in the matter herself. When we turn to agricultural field trials for different fertilizers for a given type of grain, or different types of grain for a given climate, or a combination of the two, the “patient,” the earth and her products, is taken to be blind, but we still hold it to be good practice to assign treatments to plots of land at random. Moreover, where it is practicable (e.g., with two types of fertilizer, as opposed to two types of grain) we think it would be good if the experimenter were blind, not knowing which plots had been treated in which ways.

There are several elements to the ideal empirical experiment that I have described. First, there is a notion of controls, untreated plots or persons. Second, there are the notions of blindness and double-blindness. Third, there is a notion of random assignment of treatment to a subset of the plots or persons, leaving the rest as controls. Without specifying the physical device used to apply treatments “at random,” nor other elements of the design, I shall speak of an experiment using randomization in this way as involving a randomized design.

There is a related but distinguishable idea of (random) representative sampling. Here we have a population of interest. We wish to draw an inference about its age structure, the distribution of its tastes in pizza, or whatever. We draw a representative sample. One way to do this is to draw members at random, say by assigning a number to each member of the population and then using a table of random sampling numbers to select numbers and hence members of the population.

Naturally, both random representative sampling and randomized design can be modified in all sorts of sophisticated ways. In most representative sampling of people nowadays the population is stratified into blocks with different social characteristics, and then selections are made within the strata. There are many textbooks concerning the design of randomized experiments. The first and most illustrious was R. A. Fisher’s *The Design of Experiments*, published in 1935.<sup>1</sup> This date will strike many as remarkably late. The late emergence of random representative sampling has also been remarked and studied with some care.<sup>2</sup>

<sup>1</sup> R. A. Fisher, *The Design of Experiments* (Edinburgh/London: Oliver & Boyd, 1935). Six revised editions were published during Fisher’s lifetime, and the eighth revised edition of 1966 includes posthumous revisions based upon his notes. Fisher’s even more famous *Statistical Methods for Research Workers* (Edinburgh/London: Oliver & Boyd, 1926), which went through twelve editions in his lifetime, contains considerable discussion of experimental design even in the first edition. His first published statement of the centrality of randomization in experimental design is “The Arrangement of Field Experiments,” *Journal of the Ministry of Agriculture*, 1926, 33:503–513.

<sup>2</sup> W. C. Kruskal and Frederick Mosteller, “Representative Sampling. I: Non-scientific Literature,” *International Statistical Review*, 1979, 47:13–24; “II: Scientific Literature, Excluding Statistics,” *ibid.*, pp. 111–127; “III: The Current Statistical Literature,” *ibid.*, pp. 245–265; and “IV: The History of the Concept in Statistics,” *ibid.*, 1980, 48:169–195.

I have spoken with caution of virtues that are popularly attributed to randomization. I say “popularly” because nowadays most people who can follow the ideas but have not thought much about them take for granted that randomization is not only virtuous in the case of empirical experimentation, but well-nigh a necessity in order for experimental results to have any validity. It comes as some surprise that the experts in statistical inference do not agree—or, rather, fall into several camps, each severely convinced of its own rectitude. This lack of agreement has been apparent from the beginning. Fisher’s views were widely contested from their inception in 1926. Although all branches of statistical reasoning have been greatly influenced by Fisher’s often highly innovative conceptions, his immediate employment, for a central part of his life, was in agricultural experimentation. Many of his coworkers, with whom he had close bonds of camaraderie and mutual respect—I think, for example, of W. S. Gosset, founder of the *t*-test and much else—hotly disagreed with him.<sup>3</sup> In 1932, when Fisher had a research student write a dissertation on randomized experimental design (at Rothamstead, but for a University of London degree), no one was willing to examine it, even though at the time Britain was still the leading center of pure and applied statistical theory.<sup>4</sup>

Fisher said, in essence, that one should divide fields into plots according to various geometric patterns and then randomly apply treatments to them. The geometric patterns were to be chosen on the basis of good sense and familiarity with agricultural practice. Gosset and a majority of traditionalists believed that “matched” or “balanced” arrangements were less subject to error, more instructive, and in general entitled one to draw firmer inferences. In a matched design, apparently equivalent items are paired; one is treated, the other not. In a balanced arrangement one decides, for example, that one half of the plots are to be treated and then selects them in a systematic way, following a geometrical rule. The rule is of course not chosen stupidly, but so as to counteract the expected characteristics of field variation—the fact that soil in one corner may be better than in another, or that moles are more plentiful on the southern border of the field, and so forth.

Undoubtedly Fisher won the day, at least for the following generation, but then a new, although not completely unrelated, challenge to randomized design arose. This came from the revival of the “Bayesian” school, typically associated with L. J. Savage’s theory of what he called personal probability.<sup>5</sup> Here the object is to form an initial assessment of one’s personal beliefs about a subject and to modify them in the light of experience and a theoretical analysis formally modeled by the calculus of probability and a theory of personal utility. It is

<sup>3</sup> W. S. Gosset (“Student”), *“Student”’s Collected Papers*, ed. E. S. Pearson and J. Wishart (Cambridge: Cambridge Univ. Press, 1942). On plot arrangements in field trials, see especially papers 5 (1911), 11 (1923), 13 (1926). Selection A reprints Gosset’s letters to *Nature* (1930–1936) on plot assignments. In paper 20 (1936) Gosset “ventured to point out that advantages of artificial randomization are usually offset by an increased error when compared with balanced randomization. Prof. Fisher has written a paper to test the difference of opinion between us” (p. 198). The paper in question is S. Barbacki and R. A. Fisher, “A Test of the Supposed Precision of Systematic Arrangements,” *Annals of Eugenics*, 1936, 7:189–193.

<sup>4</sup> Joan Fisher Box, *R. A. Fisher: The Life of a Scientist* (New York: Wiley, 1978), p. 150.

<sup>5</sup> L. J. Savage, *The Foundations of Statistical Inference* (London: Methuen; New York: Wiley, 1962). The criticism of randomization found here is implicit in Savage, *The Foundations of Statistics* (New York: Wiley, 1954).

widely held to be an almost immediate consequence of this approach that randomization is of no value at all (except perhaps to eliminate some kinds of fraud).<sup>6</sup> Especially in recent years, Bayesians have defended this conclusion by many formal and informal stratagems. It remains true, however, that the broad mass of routine empirical experimenters take randomized design for granted and suppose that their employers would fire them if they did not. At a more theoretical and rarefied level the randomization-is-essential versus randomization-is-irrelevant debate has inevitably become vastly more subtle than in this introductory exposition.

## II. A LACK OF PREHISTORY

My question is not about the soundness of artificial randomizers as tools for experiment and inference, but rather about their entry into experimental practice. Randomization now seems so natural that we think that it ought to have been with us since the advent of probability arithmetic and “the experimental method” in the mid-seventeenth century. Hence it comes as some surprise to learn that randomization in experiment came into common use only in the 1930s and that its point of origin was the vigorous campaign of an experimental officer at an agricultural research station in England.

The use of randomizers for making choices is ancient. Many early societies speak of the lot as a way of evading individual responsibility. A typical example, neither ancient nor recent, is the mass suicide at the mountain fortress of Masada in the Judaeen desert. In A.D. 73, when the rebellious sect of Sicarii was about to be overcome by Roman troops, the men first murdered their wives and children and then drew lots to select ten who would kill the rest. After their task was completed, they drew lots to choose one man who would kill the other nine and then himself. Josephus, who tells this story, had been in the same situation as general at Jotopata in A.D. 67. His men insisted on a mass suicide pact. They drew lots, planning that the man to draw the  $n + 1$ st lot would slit the throat of the man who drew the  $n$ th lot. Josephus was one of the last two, and he and the other survivor called it quits. In one extant version of the story, his luck was “put down to divine providence.” In another, Josephus “counted the numbers cunningly, and so managed to deceive others.”<sup>7</sup>

Every reader will have an example of the early use of artificial randomizers, and many will have examples of probable cheating, as evidenced by the story of Josephus. That cheating with randomizers is all too easy in experimental design will become evident in what follows. But there is a great distance between the use of the lot and the design of experiments. Where should we look for the beginning of randomized design?

In “empirical” experimentation, to be sure. Fisher started with field trials, a

<sup>6</sup> For a popular exposition of parts of the critique see D. A. Harville, “Experimental Randomization: Who Needs It?” *American Statistician*, 1975, 29:27–31. The conclusion has also been contested, e.g., by D. B. Rubin, “Bayesian Inference for Causal Effects: The Role of Randomization,” *Annals of Statistics*, 1978, 6:34–58; D. Basu, “Randomization Analysis of Experimental Data: The Fisher Randomization Test,” *Journal of the American Statistical Association*, 1980, 75:575–595; and Zeno Swijtink, “A Bayesian Justification of Experimental Randomization” *PSA*, 1982, pp. 124–136.

<sup>7</sup> Josephus, *The Jewish War*, trans. G. A. Williamson and E. Mary Smallwood, rev. ed. (Harmondsworth, Middlesex: Penguin, 1981); for Masada, p. 404; for the standard version of events at Jotopata, p. 220; for the less flattering Slavonic version, p. 470.

classic instance of empirical experimentation. Comparative experiments in agriculture are old. In Britain their first systematic use coincides with the agricultural reform movement. Arthur Young's *Annals of Agriculture*, commenced in 1784, gave detailed instructions.<sup>8</sup> There is ample evidence that reforming landowners paid heed from the start.<sup>9</sup> They evolved a tradition of matched plots whose defense by Gosset was mentioned above. But there was no notion of applying treatments to plots at random.

Medicine is another plausible place to look. Statistical comparisons of the quality of care or the rate of cure are quite common in the nineteenth-century literature, but seldom is there much statistical know-how and never, I think, any assignment of patients at random.<sup>10</sup>

### III. PSYCHOPHYSICS: PEIRCE'S WORK

Experimental psychology is a more promising place to look for early uses of artificial randomizers in the design of experiments. The relationships between Gustav Theodor Fechner's psychophysics and probability have been well described by Stephen Stigler.<sup>11</sup> From the 1850s Fechner was modeling sensory discriminations by the Gaussian law of error. He investigated his own ability to distinguish between very slightly different weights, but he had no assistant and was his own informant. He was not blind but knew in advance which of his boxes contained the heavier weight. He thought he could control for any unconscious bias. It was not until a quarter century later than an investigator presented weights to a blind assistant and used randomization as a control on subconscious cuing. Charles Sanders Peirce and his student Joseph Jastrow conducted such a sequence of experiments between 10 December 1883 and 7 April 1884, presenting

<sup>8</sup> The *Annals* continued for forty-five volumes. See also Arthur Young, *A Course of Experimental Agriculture* (London, 1770). Cf. W. G. Cochran, "Early Development of Techniques in Comparative Experimentation," in *On the History of Statistics and Probability*, ed. D. B. Owen (New York/Basel: Dekker, 1976), pp. 1-26.

<sup>9</sup> See, for a particular case, *Communications to the Right Honourable Sir John Sinclair, Bart., on the Subject of Drilling* (London, 1819). Drilling was a new mode of sowing, and Sinclair was the first president of the Board of Agriculture, of which Young was the first secretary. The most famous nineteenth-century work on comparative trials, with a thorough literature survey, is J. F. W. Johnston, *Experimental Agriculture, Being the Results of Past and Suggestions for Future Experiments in Scientific and Practical Agriculture* (Edinburgh: Blackwood, 1849).

<sup>10</sup> See, e.g., Jean Civiale, *Traité de l'affection calculeuse, suivis d'un essai statistique sur cette maladie* (Paris, 1838). The statistical essay occupies pp. 548-699 and had been submitted for the Prix Montyon in 1833. See the report by Simeon-Denis Poisson, Pierre-Louis Dulong, Jean-Dominique Larrey, and François-Joseph Double, *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 1835, 1:167-177. At issue was a comparison between two (highly disagreeable) operations for gallstone. For the traditional method Civiale reports 5,615 operations, after which 1,141 patients died; while with the new operation there were 307 patients, with 6 deaths and 3 undetermined outcomes. The referees praise Civiale's work and contrast it with most statistical studies in medicine, which consist of facts "without authenticity, without detail, without control, without value." (p. 169) Despite their respect for Civiale's work, they assert that "in practical medicine facts are too few in number to enter into the domain of the calculus of probabilities" (p. 174), and hold that the law of large numbers—Poisson had invented the term that very year, 1835—could not apply. For another tradition in experimental medicine see William Coleman, "Experimental Physiology and Statistical Inference: The Therapeutic Trial in Nineteenth-Century Germany," in *The Probabilistic Revolution*, Vol. II: *Ideas in the Sciences*, ed. L. Krüger, G. Gigerenzer, and M. S. Morgan (Cambridge, Mass.: MIT Press, 1987) pp. 201-226.

<sup>11</sup> Stephen M. Stigler, *The History of Statistics* (Cambridge, Mass.: Harvard Univ. Press, 1986), pp. 239-261.



their conclusions to the National Academy of Sciences on 17 October. Stigler writes that "the Peirce-Jastrow experiment is the first of which I am aware where the experimentation was performed according to a precise, mathematically sound randomization scheme!" He says that "as a nineteenth-century experiment it was unexcelled" and would compare well with present-day work in the psychological laboratory.<sup>12</sup>

Peirce was a long-time student of errors of observation in geodesy and astronomy and had refined the theory of error to describe experimental phenomena in those domains.<sup>13</sup> The work of Fechner and his successors in psychophysics was well known. Fechner held that there is a threshold below which one cannot discern small differences in, for example, weight. Peirce set out to refute this, holding that there is instead a continuum of decreasing accuracy adequately described by the law of error.

In the experiment Peirce conducted with Jastrow the experimenter or operator was separated from the subject by a screen, through which an adapted post office balance was inserted. On the experimenter's side the pan had a 1 kg weight on it; in addition, a small weight could be added to or taken off the pan. On the other side the pan was brought up to the point where it exerted pressure on the subject's finger. The experimenter presented weights to the subject in two ways. In the first, the subject was given the pressure of the 1 kg weight to experience; then the small weight was added after the subject said "change"; finally, after saying "change" again, the subject was presented with the pressure caused by the 1 kg weight. In the other case the subject was first presented with the pressure of 1 kg plus a small extra; then the 1 kg weight; and then, once again, 1 kg plus a small extra. Great mechanical care was taken to ensure that there would be no cue, except felt pressure, as to whether a weight was being added or removed. In addition the order of the trials was determined at random by a pack of playing cards. If red was drawn, the experimenter presented first the 1 kg weight, then a supplemented weight, and then the original weight. But if black was drawn, the order of presentation was reversed.

Aside from the mechanical devices employed in the presentation of the weights, the design of the experiment employed three novelties: the randomization was deliberate; the subject was not allowed to say that he could not decide whether the weight had been increased or decreased between the first and second presentations, and the subject reported confidence in his judgment on a scale of 0 to 3.

Peirce noted that randomization prevented the subject from guessing the order in which the weights would be presented. There would, however, be occasional long runs of one or the other color by chance "which would tend to confuse the mind of the subject. But it seems clear that this disadvantage was less than if any means had been taken to prevent them."<sup>14</sup> This issue of confusing long runs has

<sup>12</sup> Charles S. Peirce and Joseph Jastrow, "On Small Differences of Sensation," *Memoirs of the National Academy of Sciences*, 1885, 3:75–83; and Stephen M. Stigler, "Mathematical Statistics in the Early States," *Ann. Statist.*, 1978, 6:239–265, on p. 248.

<sup>13</sup> Charles S. Peirce, "On the Theory of Errors of Observation," Appendix 21 of *Report of the Superintendent of the U.S. Coast Survey for the Year Ending June 1870* (Washington, D.C.: U.S. Government Printing Office, 1873), pp. 200–224 and plate 27. This remains one of the most detailed empirical studies of the law of errors of observation.

<sup>14</sup> Peirce and Jastrow, "Small Differences of Sensation" (cit. n. 12), p. 80.

been a bone of contention ever since Fisher reintroduced randomization in experiment. In fact Peirce was at pains to exclude excessively long runs. An experiment consisted of a block of fifty trials, and fifty playing cards were used to randomize. But instead of using a shuffled pack of fifty cards Peirce used two packs, one of thirteen red, twelve black, and one of thirteen black, twelve red, running through one before the other. (One supposes that the canny Peirce used a randomizer such as a coin to choose which of the two subpacks would commence the experiment.)

Peirce was altogether clear on his use of randomizers. He was less explicit about the mathematical analysis of his data than about the mode of its collection, but he appears to have incorporated the mathematical properties of his device into the inferences that he drew from his experiments. Yet in the ensuing literature randomization entirely dropped out. In their classic 1892 study of small differences in sensation, G. S. Fullerton and James Cattell would merely write, for example, "Of course the different experiments in each series were made in an irregular order and the observer had no clue to guide him except the objective difference in the lights."<sup>15</sup> (These authors tested discrimination for a wide range of phenomena, including very similar lights.) They might of course have used an artificial randomizer to ensure irregularity, but they do not seem to consider how it matters or that there may be any conceptual difficulty in the concept of irregularity.

Although Jastrow provided a standard exposition of the Peirce-Jastrow procedure in 1888, he said nothing of randomization.<sup>16</sup> Despite the later work of investigators such as Jastrow, Fullerton, and Cattell, Peirce's research was ill-received. One reason was the second novelty mentioned above: Peirce did not allow his subject the option of being undecided as to whether a weight had been increased or decreased. This was held to be unfaithful to psychological reality. E. B. Titchener provided a thorough survey of complaints of this sort, concluding with his own scathing remarks on Peirce's idea.<sup>17</sup> Throughout the debate the procedure of randomization is just never mentioned. Naturally it was important to Peirce's program to insist that the subject should always cast a vote, even when he had no confidence in the judgment expressed. For Peirce's thesis was that as two stimuli become more and more similar, people become continuously less good at telling them apart, but there is no threshold below which judgment is entirely worthless.

It was precisely because of this thesis that Peirce introduced his third novelty, according to which the subject ranked his confidence in his judgment on a scale of 0 to 3. Why 0 to 3? Ever practical, Peirce marked each playing card, starting in the upper left corner, with no punches, one punch, two punches, or three punches. The experimenter then kept score as follows. If the subject judged correctly, the card was placed face down; if incorrectly, the card was placed face up. Moreover, the card was placed with no, one, two, or three holes in the upper left corner of the pile, according as the expressed degree of confidence was 0, 1,

<sup>15</sup> G. S. Fullerton and James McKean Cattell, *On the Perception of Small Differences* (Publications of the Univ. Pennsylvania Philosophical Society, 2) (Philadelphia, 1892), p. 237.

<sup>16</sup> J. J. Jastrow, "A Critique of Psycho-Physic Methods," *American Journal of Psychology*, 1888, 1:271-309.

<sup>17</sup> E. B. Titchener, *Experimental Psychology: A Manual of Laboratory Practice*. Vol. II: *Quantitative Experiments*. Pt. 2: *Instructor's Manual* (New York: Macmillan, 1905), pp. 285-291.



2, or 3. This enabled accurate scorekeeping to be done as quickly as possible, to avoid tiring or boring the subject. Peirce expected that the subject's confidence would decline as the objective probability of correctness of his guesses decreased, and he derived a logarithmic formula to this effect that roughly fit his data. Here his background interest was that of the practical geodesist.<sup>18</sup>

Comparative testing of subjects that were blind as regards the treatments administered became standard in psychology. Randomization had to wait three decades and even then was regarded as a trifling adjunct to an experiment. The features that interest a statistician went unexamined. As for Peirce, he was convinced that his experiments had shown that there is no minimum perceptible difference. His concluding sentences are instructive. It is to be recalled that 1883 was the year of his painful divorce and remarriage, a divorce that hinged in part on what a woman might be expected to know, and that 1884 was the year that Peirce's appointment at Johns Hopkins was terminated.

The general fact [of no threshold] has highly important practical bearings, since it gives new reason for believing that we gather what is passing in one another's minds in large measure from sensations so faint that we are not fully aware of them, and can give no account of how we reach our conclusions from such matters. The insight of females as well as certain "telepathic" phenomena may be explained in this way. Such faint sensations ought to be fully studied by the psychologist and assiduously cultivated by every man.<sup>19</sup>

The word *telepathy* was less than two years old when Peirce wrote these words, but the Eastern seaboard was awash with enthusiasm for the idea. The American Society for Psychical Research was about to be founded, and its first project was a survey to determine the frequency of telepathic phenomena.<sup>20</sup>

It is hardly surprising that, having turned to psychology, we should now move to an inquiry on its margins, one that has been called parapsychology. The Gossets who preferred balanced design were not mere traditionalists. They had good arguments (which, indeed, to an experimenter are more compelling than the purely abstract schematism of the Bayesian latter-day antirandomizers). The arguments were in part based on experience of the causes of variability in field plots and systematic ways to avoid those causes. But suppose we turn to a subject about which nothing whatsoever is known, so that one cannot even begin to speculate intelligently about causes of variation? Suppose, further, that it is a subject redolent of fraud, an activity that even Bayesians admit can be controlled for by randomization? That is where we should be looking: psychic research. The more "empirical" and nontheoretical a question, the more randomization makes sense.

<sup>18</sup> Stigler notes that although Peirce was philosophically an "objective frequentist," he was perhaps the first experimenter to "elicit subjective or personal probabilities, determining that these probabilities varied approximately linearly with the log odds": "Mathematical Statistics" (cit. n. 12), p. 248. What Peirce himself wrote is that "what we have called the degree of confidence was probably the secondary sensation of a difference between the primary sensations compared": Peirce and Jastrow, "Small Differences of Sensation" (cit. n. 12), p. 82. The log odds is  $\log(p/1 - p)$ , where  $p$  is the objective probability of giving a correct judgment. For interesting comments on Peirce and the log odds see the most forceful present-day advocate of the log odds as a tool in inference: I. J. Good, *Good Thinking* (Minneapolis: Univ. Minnesota Press, 1983); see the index under "Peirce."

<sup>19</sup> Peirce and Jastrow, "Small Differences of Sensation," p. 83.

<sup>20</sup> The formal organizational meetings were 23 Sept. and 18 Dec. 1884, just before and just after Peirce and Jastrow presented their paper.

## IV. TELEPATHY AND THE SOCIETY FOR PSYCHICAL RESEARCH

Most of us now tend to lump telepathy together with spiritualism, or, better, what I shall call spiritism, since the former also denotes a more or less coherent body of religious beliefs. Spiritism includes beliefs in poltergeists, ectoplasm, and the various activities of spirits that are provoked by the intercession of mediums. We now tend to put telepathy in with all that. In fact telepathy was a scientific reproach to spiritism.

Spiritism and mediums came into the world in 1848 in Hydesville, near Arcadia, New York. The precise site was the home of Mr. and Mrs. John Fox, who heard knocking. The fad took America and Europe by storm. As Henri Ellenberger states: "At the beginning of 1852 the wave of spiritism crossed the Atlantic, invading England and Germany. In April 1853 it swept over France and soon reached all parts of the world."<sup>21</sup> The *Oxford English Dictionary* reports that the expression *thought-reading*—the spiritist term for which *telepathy* was proposed as a "scientific" substitute, was used in print in 1850. The word *medium* in the pertinent sense came in at about the same time.

Although mediums had various pretensions, their chief claims were three in number. First, they were able to produce remarkable phenomena in certain circumstances. Second, they were able to know a great deal about the past and present of members of the séance, even those whom they had never set eyes on before. Third, they were in contact with members of the spirit world—souls (or whatever) of the departed—who produced the phenomena and the information. A substantial number of what I shall call scientific persons thought, often from personal experience, that the first two claims were substantiated, despite the number of known frauds. But they proposed an alternative hypothesis for the third claim, in particular as it concerned knowledge about other people. The medium's information did not necessarily come from the spirit world. It was obtained by transference of thoughts, or telepathy, from other members of the audience.

Several scientifically minded societies were formed to investigate the phenomena. The first and most important was the Society for Psychical Research, founded in London in 1882. Its members and associates included celebrated physical scientists of the day: Sir William Crookes, Lord Rayleigh, and Sir Oliver Lodge, for example. Crooke's radiometer—small painted vanes in a vacuum that rotate in sunlight—was invented as a tool for investigating what has since been called psychokinesis (minds making motion). Its physics still happens to be a little in question, with the pressure of light competing with convection currents as the conjectured cause of rotation. The leading scientists who studied telepathy were drawn from the physical rather than the life sciences. Several reasons for this are readily perceived. Inexplicable emanations and vibrations were attractive to those engrossed with the problems of the electromagnetic ether. In contrast, Alfred Russel Wallace, the naturalist and evolutionary pioneer, parted from the SPR on the grounds that its standards of evidence and models of reality, derived from physics, were inappropriate. He remained an "honourary member," but preferred to do his psychic research outside of the society, with mediums.

The men and women of the Society for Psychical Research came from an

<sup>21</sup> Henri Ellenberger, *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychology* (New York: Basic Books, 1970), p. 84.

intellectual elite whose center was Cambridge University. They found the raging popular middle-class spiritualism and spiritism entirely repugnant.<sup>22</sup> Their scientific approach was an attempt to preserve the phenomena for purer minds. Crookes, Lodge, and Rayleigh did embrace a deep-seated spiritualism to which they were far more committed than the chiefly curious nonscientist intellectual members of the society. They did not want spiritualism debased by fads and enthusiasms, and they preferred not to resort to mediums or other "sensitives" for their phenomena.

Henry Sidgwick, Knightbridge Professor of Moral Philosophy at Cambridge, was first president of the SPR. His most lasting written work is his *Methods of Ethics* (1874), which is still a subject of study today. He was an important public figure and the leading man of his day to work for the higher education of women. He and Mrs. Sidgwick (E. M. Balfour) were responsible for establishing the first women's college near the center of Cambridge University. (Girton College had already been established in a village outside Cambridge.) Incidentally, Sidgwick was only the first British philosopher to be intimately connected with psychic research. The tradition continued with H. H. Price and C. D. Broad and died out only very recently. The attitude of the philosophers, from Sidgwick on, was always skeptical interest. Most psychic phenomena, they thought, were hoaxes, frauds, wishful thinking, or whatever, but there remained a residue of inexplicable well-attested facts that could not be laid aside just because they were so discordant with the present scientific world view.

The word *telepathy* is coeval with the society, having been coined by one of its founding members, F. W. H. Myers, in 1882.<sup>23</sup> For the first several years the SPR was preoccupied with telepathy. It remained a prime concern for decades. The 1911 article in the *Encyclopaedia Britannica* is a good state-of-the-art exposition, written by a member of the SPR who was, atypically, an anthropologist. It notes that "as used by the society the term 'telepathy' is a mere designation, and implies no hypothesis as to 'action at a distance' or the operation of any force not recognized by physical science."<sup>24</sup> *Telepathy* was, then, to be a neutral term

<sup>22</sup> See Brian Wynne, "Natural Knowledge and Social Context: Cambridge Physicists and the Luminiferous Ether," in *Science in Context: Readings in the Sociology of Science*, ed. Barry Barnes and David Edge (Cambridge, Mass.: MIT Press, 1982), pp. 212–231. It is wrong to emphasize, to the exclusion of others, members such as the Sidgwicks, Wallace, Lodge, Rayleigh, or Crookes, who established great reputations in other endeavors. There are also the men who "found themselves" in the SPR. Edmund Gurney, brilliant at everything, decided in nothing, is the classic example; for an informative but speculative study see Trevor H. Hall, *The Strange Case of Edmund Gurney* (London: Duckworth, 1964). F. W. H. Myers is another who needed the SPR to find a niche in the world. For a sympathetic portrayal see William James's *eloque*, "Frederick Myers' Services to Psychology," *Proceedings of the Society for Psychical Research*, 1901, 17:13–23. In January 1902 the two men were staying in the same hotel in Rome during Myers's final days. James wrote of the event to his brother Henry, "I fear that the S.P.R., with Myers gone, has hardly any prospect of much longer continuance. Somehow I feel very tenderly towards him, and regret not being warmer with him when I had him": William James, *The Works of William James*, ed. F. H. Burkhardt et al., Vol. XVI: *Essays in Psychical Research* (Cambridge, Mass.: Harvard Univ. Press, 1986), p. 486.

A thoroughly balanced account of the SPR is given by Janet Oppenheimer, *The Other World: Spiritualism and Psychical Research in England* (Cambridge: Cambridge Univ. Press, 1985). More recent and American parts of the story are told by Seymour H. Mauskopf and Michael R. McVaugh, *The Elusive Science: Origins of Experimental Psychical Research* (Baltimore: Johns Hopkins Univ. Press, 1981).

<sup>23</sup> "Report of the Literary Committee by W. F. Barrett et al., E. Gurney and F. W. H. Myers, Hon. Secs.," *Proc. Soc. Psych. Res.*, 1882, 1:116–155, on p. 147. Myers wrote the report.

<sup>24</sup> N. W. Thomas, "Telepathy," *Encyclopaedia Britannica*, 11th ed. (1911), Vol. XXVI, p. 546.

denoting observed phenomena. As one of the earliest members of the society was to remark,

our own conception of telepathy involves, strictly speaking, no theory at all. It involves as little of pure theory as Newton's conception of gravitation. What Newton did was to find the simplest general expression for the observed facts by saying that the heavenly bodies acted upon each other with a certain measurable force. He did not attempt to explain the mode of the action; he contented himself with asserting its existence and defining its limits.<sup>25</sup>

Throughout these discussions the highest standards of positivist scientific methodology were observed. Or rather, the highest standards of theoretical methodology, the sort of thing still often taught in philosophy of science courses today. The enthusiasts of the society were not, however, so attentive to their experiments. Quite aside from charges of fraud, they were not able to think very carefully about how to analyze their data.

#### V. RICHEL'S RANDOMIZERS

There were a number of standard topics for investigation in the early days of the Society for Psychical Research. They are still the stock in trade of psychics and magicians. Objects were hidden in a part of a room, and people who did not know what they were, were asked to draw pictures of them. One person would form an image, which the other members of the party would sketch. Cards were drawn and sealed in envelopes; then others would guess the card. Words were written down and their successive letters guessed. These early inquiries of the SPR needed no mediums, no enthusiasts, no sensitives. They were "objective." They required no admixture of social classes. The society also examined reports of phenomena from outside its circle of members and friends, but it found its investigations of telepathy peculiarly fitting to its status and associated methodology.

Nevertheless, the first methodological breakthrough came from across the Channel, from Charles Richet. Richet was a Parisian physiologist with a fascination with the odd. Although the study of hypnotism had become disreputable, Richet was much taken by its vogue in stage performances and reintroduced it to scientific study, to Jean-Martin Charcot, Pierre Janet, and Sigmund Freud. It was he who most seriously took up the phenomenon of automatic writing. Whenever there was something strange connected with the psyche, Richet was the first to look into it. He deserves a far larger place in histories of the origin of psychoanalysis than is commonly granted. He took eagerly to split personalities, which manifested themselves and were fully seen as such first in 1875. Richet was so excited by the phenomenon that in 1889, under the pseudonym Charles Epheyre, he published a split-personality novel whose characters are far truer to the clinical practice of the day than any Jekyll and Hyde.<sup>26</sup>

Richet read the *Proceedings of the Society for Psychical Research* as soon as

<sup>25</sup> Frank Podmore, review of J. Ochorowicz, "De la Suggestion mentale," *Proc. Soc. Psych. Res.*, 1886, 4:562–568, on p. 568.

<sup>26</sup> Charles Epheyre [i.e., Charles F. R., or Charles F. Richet], "Soeur Marthe," *Revue des deux mondes*, 1889, 93:384–431.

they commenced in 1882. The reports on telepathy introduced him to another strange aspect of the mind, which he studied intensively from 1884 to 1890. Richet's first article on the subject, which ran to sixty-six pages, bore the title "La suggestion mentale et le calcul des probabilités."<sup>27</sup> "Suggestion mentale" is his rather positivistic choice of words for thought transference.

We now think of probability as so integral a part of scientific method and practice that we are astonished to recall how ground-breaking a paper this is. Richet was only speaking the truth when, in 1884, he wrote of "using a method which is extremely rarely in usage in the sciences, the method of probabilities."<sup>28</sup> His first run of experiments was conducted 25 June to 13 July 1884—by chance, only a few months after those of Peirce and Jastrow.

Richet hardly introduced probabilistic reasoning into psychology, let alone into science in general. Fechner, Wilhelm Wundt, and other investigators of psychophysics had been using probability modeling since the 1850s. Hermann Ebbinghaus was modeling short-term memory by probability in 1879, although he published this work only in 1885. But it remains true that, outside of astronomy and geodesy, probability had very little role in scientific inference at the time that Richet wrote. Biometrics, so valuable a source of probability ideas by 1900, and the origin of a lot of our statistical ideas, was in its infancy, or perhaps its prenatal state. Psychophysics primarily involved modeling phenomena by probability structures and was little concerned with questions of inference. Richet, although using the most trivial of probability models for cards, was concerned above all with questions of probable inference. He did not invent a single idea and was indeed quite ignorant of ideas known to Pierre-Simon de Laplace a century earlier. But he did something that, though obvious, no one had quite thought of doing before.

After reading the *Proceedings of the Society for Psychical Research* for a year or so, Richet (like many others) came to the view that weak powers of telepathy might be common to the population at large. How would one detect a weak ability of this sort? Only sometimes would a person actually transfer the thought of another to his or her own mind. So Richet proposed a long sequence of trials in which an "agent" drew a playing card at random and concentrated upon it for a short time, after which a "reagent" guessed the suit of the card. Some 2,927 guesses were made by various reagents in sequences of various length. In each sequence the expected number of successes was compared with the actual number, and a rather primitive statistical inference was drawn. In the total sequence of guesses there were 789 successes, whereas 732 is the expected number. "Among adult persons in good health who are not hypnotised, it is possible that mental suggestion [i.e., what the SPR was calling telepathy] can be experienced. This mental suggestion is even, in a small measure, probable, but with a probability that scarcely exceeds  $\frac{1}{16}$ " [a misprint for  $\frac{1}{10}$ ].<sup>29</sup> That is,

<sup>27</sup> Charles Richet, "La suggestion mentale et le calcul des probabilités," *Revue Philosophique de la France et de l'Étranger*, 1884, 18:609–674. The SPR published both the *Proceedings* and the *Journal of the Society for Psychical Research*. The former was available to the public by subscription, while the latter was intended for private circulation among members.

<sup>28</sup> Richet, "Suggestion mentale," p. 616.

<sup>29</sup> *Ibid.*, p. 632. When he turned to the ability of a few individuals to transmit drawings telepathically, his results were much more favorable: "The probability in favor of the reality of mental suggestion can be represented by  $\frac{3}{4}$ ," p. 670.



there is some probability for the hypothesis of mental suggestion—low, but not negligible.

The result was not very thrilling for the SPR, but the method was. Edmund Gurney, one of the founding members of the society, concluded in a résumé of Richet's paper that it "can hardly fail to be a permanent landmark in the slowly widening field of psychical discovery." In the paper Richet, disappointed with cards, went on to various other probabilistic studies. He conjectured that telepathy was less a mental than a motor phenomenon, so that your thoughts would show up in my muscles rather than my mind. He tested this notion probabilistically, with various kinds of divining rod. An object was buried under one of twelve trees in an orangery or hidden in a certain shelf of Richet's large library. He found better results with rods than cards; the outcome was better still when he experimented with various kinds of "lucidité" or clairvoyance. These last experiments involved sketching concealed objects. Gurney complained of a "caution of statement in the wrong place." He thought that Richet had derived too low a probability for telepathy from his data. If caution were called for, it was on the score of the data themselves. Might cheating or cuing have been involved?<sup>30</sup> In the next section I return to the issue of how to assess the probabilities given the data—not one of Gurney's strong points.

In 1888 Richet published a further 150 pages in the *Proceedings* of the SPR, followed in 1890 by a curious paper on "hypnotic lucidity or clairvoyance." He had decided that the work with cards showed nothing. Moreover, this very fact cast doubt on experiments on clairvoyance, in which one person was supposed to reproduce a diagram shown to another. In the case of cards, "we have a fact absolutely negative, which must inevitably cast some doubt on the experiments in the reproduction of diagrams."<sup>31</sup> He had become convinced that clairvoyance could exist only in exceptional persons. This was an era, especially in France, of great popular interest in somnambulism, hypnotism, automatic writing, hysteric anesthesia, and the like. That interest would furnish the unusual people Richet needed. Richet began exhausting experiments on a woman whom Janet had already made well known; we might speak of Janet loaning her to Richet. One playing card was placed in a sealed envelope; over several hours, she had to bring the card to her mind. Richet's dedication was matched only by his incompetence. The woman in question, one Léonie or Léontine, was one of the famous early split personalities (she had a third personality who apparently did not make herself available to Richet).<sup>32</sup> She had long been a stage companion of conjurers and mesmerists, and one of her personalities was a great tease. It is so obvious that she would have taken Richet in—given his own accounts of his

<sup>30</sup> Edmund Gurney, "M. Richet's Recent Researches in Thought-Transference," *Proc. Soc. Psych. Res.*, 1884, 2:239–257, on p. 257.

<sup>31</sup> Charles Richet, *Rélation de diverses expériences sur la transmission mentales, la lucidité, et autres phénomènes non explicables par les données scientifique actuelles*, *Proc. Soc. Psych. Res.*, 1888, 5:18–168; quoting from p. 149; repeated in Richet, "Further Experiments in Hypnotic Lucidity or Clairvoyance," *ibid.*, 1890, 6:66–83, on p. 66.

<sup>32</sup> Ian Hacking, "The Invention of Split Personalities," in *Human Nature and Natural Knowledge*, ed. Alan Donagan, Anthony N. Perovich, Jr., and Michael V. Wedin (Boston/Dordrecht: Reidel, 1986), pp. 63–85. From 1875 multiple personalities were amply studied by the French psychologists, and by 1890 they were becoming common in America. Although there were no English splits, the idea of multiples fascinated the members of the SPR, whose *Journal* and *Proceedings* discussed the French and American cases at length.

“experiments”—that one blushes for the good man. He himself noted that the envelopes were only opaque but persuaded himself that Léonie was not cheating.

So much is an aside. Richet introduced the study of probability into psychic research—and as an inferential tool into psychology in general—by testing the ability of reagents to improve on chance by guessing at randomly drawn cards. Second, in these experiments he drew entirely negative conclusions in the end. This is not to say that his enthusiasm for psychical enquiries came to an end. On the contrary: he became convinced only that probability was the wrong investigative tool. One should not use randomizers on a large and typical population, but rather should turn one’s attention to sensitives.

Richet’s career as a physiologist flourished, being crowned with a Nobel Prize in 1913 for his research on the causes of certain sudden deaths, which proved to be due to an unhappy combination of antigens. (He named this anaphylaxis.) Yet he was utterly faithful to the discipline for which he invented the French name of *métapsychique*, a word that the reader may have to examine more than once. He established an *Annales des Sciences Psychiques* in 1891 and an international society, of which he was president, in 1930, when he was eighty years of age. His *Traité de métapsychique* (1922) is dedicated to the memory of the great pioneers Gurney and Myers.<sup>33</sup> His physiological sense prevented him from entertaining the hypothesis of spiritual survival after death. There is also evidence that he was unimpressed by the banal and inane character of communications from the departed—not that he was ever as frank as T. H. Huxley, who remarked: “The only good that I can see in the demonstration of the truth of ‘Spiritualism’ is to furnish an additional argument against suicide. Better live a crossing-sweeper than die and be made to talk twaddle by a ‘medium’ hired at a guinea a *séance*.”<sup>34</sup>

## VI. EDGEWORTH ON THE ANALYSIS OF RICHEL’S DATA

Gurney summarized Richet’s work in the *Proceedings* of the SPR. The issue is dated 30 December 1884. Before it went into print the presses were stopped to allow the inclusion of a letter from Oliver and Alfred Lodge, dated 8 January 1885. It tried to provide a method for analyzing Richet’s data that improved on Richet’s own. It offered what I would call a Bernoullian (or roughly confidence-interval) technique in terms of error probabilities. Gurney retorted in a further appendix with a Bayesian analysis. Both were primitive, but both were written on the spur of the moment by men whose expertise was not probability. An expert had to be called in, and the best one was readily available in the person of the economist F. Y. Edgeworth. He had just written a general analysis of statistical inference, and he provided two papers for the *Proceedings*, in 1885 and 1887.<sup>35</sup> These are fine papers, beautiful enough almost to justify the entire subject of parapsychology, for which Edgeworth had little use.

<sup>33</sup> Charles Richet, *Traité de métapsychique* (Paris: Alcan, 1922).

<sup>34</sup> Leonard Huxley, ed., *Life and Letters of Thomas Henry Huxley*, 2 vols. (New York: Appleton, 1900), Vol. I, p. 452. Quoted in William James, “Confidences of a Psychical Researcher,” *American Magazine* (Oct. 1909); reprinted as “Final Impressions of a Psychical Researcher,” in James, *Memoirs and Attitudes* (London: Longmans, Green, 1911), p. 186.

<sup>35</sup> The Lodges’ letter is printed in *Soc. Psych. Res.*, 1884, 2:257–262; Gurney’s reply follows, pp. 262–264. Edgeworth’s papers appeared later: F. Y. Edgeworth, “The Calculus of Probabilities Applied to Psychical Research,” *Proc. Soc. Psych. Res.*, 1885, 3:190–199; and Edgeworth, “The Cal-

The quality of the papers will be more apparent to the probabilist than to the general historian, and I shall not report them in detail. In part he plays Solomon between Gurney and Lodge, but without mentioning them by name. He first of all points out that the experimenter must choose the physical model in terms of which to test the data to see whether they indicate the presence of some anomaly. He speaks of "two schemata that present themselves, each recommended by high authority." In the first we proceed "by inverse probability, upon the principle of Bayes." After developing a formula he continues with the second procedure: "But however well established the preceding formula as an organon of statistics, the following schema, savouring more of Bernoulli than of Bayes, is perhaps more appropriate to the particular problem at hand."<sup>36</sup>

He does conclude that the probability that Richet's phenomena were obtained by chance is very small, 0.00004, and so the reliability of the phenomena *not being due to chance* "may fairly be regarded as physical certainty" and "the conclusion may be regarded as safe." After this, with some sagacity, he concludes his 1885 paper with the words: "Such is the evidence which the calculus of probabilities affords as to the existence of an agency other than mere chance. The calculus is silent as to the nature of that agency—whether it is more likely to be vulgar illusion or extraordinary law. That is a question to be decided, not by formulae and figures, but by general philosophy and common sense." This warning may be compared with a rather celebrated assertion by R. A. Fisher on the logic of significance testing. He speaks of a "logical disjunction" being the basis of a test of significance. Either something very uncommon has occurred by chance, *or* a hypothesis of "no effect" must be rejected. Fisher explicitly introduced these observations in connection with "the studies known as parapsychology."<sup>37</sup>

## VII. LEHMANN'S CRITIQUE: "NUMBER HABITS"

Richet's "Suggestion mentale" attracted interest far beyond the confines of the SPR. Throughout 1885 the Russian medical journal *Vratch* summarized his investigations.<sup>38</sup> Experiments were conducted to prove that his evidence for telepathy—evidence that Richet himself was to abandon—could be simulated by chance devices such as the results of national lotteries. Wilhelm Thierry Preyer in Leipzig was perhaps the most instructive of these critics. Ludwig Kuhlenbeck, writing in the new German theosophical and spiritualist journal *Sphinx*, may have been the cruelest.<sup>39</sup>

Somewhat later, from Copenhagen, F. C. C. Hansen and Alfred Lehmann

---

culus of Probabilities Applied to Psychical Research, II," *ibid.*, 1887, 4:189–208. For Edgeworth's twelve other papers on the foundations of probable inference and the theory of error, composed in the same time frame (1883–1887), see the bibliography in Stigler, *History of Statistics* (cit. n. 11), p. 379f.

<sup>36</sup> Edgeworth, "Calculus of Probabilities [I]," p. 191.

<sup>37</sup> *Ibid.*, p. 199; and R. A. Fisher, *Statistical Methods and Scientific Inference* (Edinburgh/London: Oliver & Boyd, 1956), p. 43.

<sup>38</sup> "M. Gomalès on the Experiments of M. Charles Richet, *Vratch* 9/21 May, 16/28 May," *Journal of the Society for Psychical Research*, 1885, 2:33.

<sup>39</sup> W. T. Preyer, *Die Erklärung des Gedankenlesens* (Leipzig, 1886). For Preyer's views on the SPR see "Telepathie und Geisteshererei in England," *Deutsche Rundschau*, Jan. 1886, 2:30–51. Gurney characteristically replied with a sixty-three page pamphlet: *Telepathie: Eine Erwiderung auf die Kritik der Herrn. W. Preyer* (Leipzig, 1887). L. Kuhlenbeck, "Die sogenannten spiritischen Versuche des Professors Charles Richet," *Sphinx*, Sept. 1888, 6:172–183, on p. 177.

published other and more detailed conjectural explanations of the phenomenon of thought transference. They surveyed the literature, chiefly as furnished by the SPR; their main target, however, was a series of experiments conducted by the Sidgwicks, who used a young man who was probably a scoundrel, G. A. Smith, and four youths whom he had hypnotized. These experiments involved not cards but the transfer of numerals from one mind to another. Hence one issue was the initial selection of the digits. In addition there were the usual questions about possible cuing, for which Hansen had an ingenious explanation, and fraud.<sup>40</sup>

Incidents involving Smith are among the least delectable in the history of the SPR, and it has been conjectured that Gurney committed suicide on learning that Smith was a fraud.<sup>41</sup> But Hansen and Lehmann did not know the inside story, and dealt only with the data as reported. Lehmann and Hansen proposed that two unlooked-for effects of this type of experiment jointly produce positive results. Their first paper appeared under the title "On Involuntary Whispering," which referred to Hansen's contribution. He argued that at least from the time of Richet experimenters had unwittingly whispered the name of the card or number they were thinking of, thus cuing the subject by unnoticed but indisputably physiological means. Hansen's devices to test and illustrate this idea were ingenious, but his hypothesis does not concern us here. Lehmann's contribution is the one that bears on randomization.

Lehmann remarked that when people guess cards or numbers they make some guesses more frequently than others. The observation was not novel. In 1886 Charles Sedgwick Minot, reporting to the American Society for Psychical Research, had noticed that a great many people have what he called "number habits." Surveying chiefly the work of the SPR in London, he concluded that this fact is "unfavorable to the ideas of thought-transference having taken place in any of the experiments with digits." The reasoning was that the digits guessed by a subject were not "caused" to occur in their order by transference of thoughts from another person. On the contrary, the cause was a "characteristic and personal number-habit, and this habit regulates" the guesses. The habit in some cases is shown to be "exceptionally rigid and persistent."<sup>42</sup> It leaves no room for action at a distance by the thoughts of another.

Lehmann made his observation to support a different point. If both the experimenter and the subject had the same or similar number habits, then the numbers they produced would show much better than merely chance agreement. This possibility, argued Lehmann, vitiated the experiments performed with Smith as the main agent.

Lehmann's critique led to a spirited five-year debate involving William James, E. B. Titchener, Henry Sidgwick, and others. It concluded with an unsigned piece in the *Journal of the Society for Psychical Research* that recapitulated correspondence already published there, with some editorial remarks that are almost certainly by F. W. H. Myers. The author asserts: "As early as 1886,

<sup>40</sup> F. C. C. Hansen and A. Lehmann, "Ueber unwillkürliches Flüstern: Eine kritischen und experimentelle Untersuchung der sogenannten Gedanken-Übertragung," *Philosophische Studien*, 1895, 11:471–530. See also Prof. and Mrs. H. Sidgwick, "Experiments in Thought Transference," *Proc. Soc. Psych. Res.*, 1889/90, 6:128–170.

<sup>41</sup> See Hall, *Edmund Gurney* (cit. n. 22). Hall discounts a possible sexual liaison with Smith's curious band of young men as a cause of the sudden and "unexplained" death of Gurney.

<sup>42</sup> Charles Sedgwick Minot, "The Number Habit," *Proceedings of the American Society for Psychical Research*, 1885–1889, 1:86–95, on pp. 94–95. The issue is dated July 1886.

therefore (see *Phantasms of the Living*, Vol. I., pp. 31–35, and Vol. II, p. 653), experimenters who worked in connection with the Society for Psychical Research were accustomed to use the method of drawing numbers at random, and it is hardly necessary to say that all the number-guessing in the experiments of Professor and Mrs. Sidgwick was carried out on this plan.”<sup>43</sup> Alas, matters are not quite so simple. It *was* necessary to say that the Sidgwicks had used the plan of random sampling numbers. They seem not to have said so themselves, although given the discussion initiated by Richet’s paper of 1884, one would expect the idea to have been commonplace.

As for the reference to *Phantasms of the Living*: this is the work of Gurney, Myers, and Frank Podmore, stalwarts of the SPR.<sup>44</sup> It was published in 1886, and its main project was to conduct a “census” of persons who had hallucinated the appearance of a person within twelve hours of the latter’s death, and with no grounds for believing that the person would die at that time. These events were found to be sufficiently common in a sample to argue that these ghost stories (as Peirce was to call them derisively) were strong evidence of a certain type of telepathic communication. The labors of the three men (especially Gurney) were prodigious in this enterprise, to which I shall return. But there were also passing references to other evidences for telepathy, and the passage cited first by Myers (Vol. I, pp. 31–35) considers the work of Richet. However, the data supporting the claim that from 1886 SPR members always randomized were excessively thin.

On page 34 of Volume I there is exactly one reference to randomizing: “And still more remarkable is the result obtained by the Misses Wingfield, of the Redings, Totteridge, in some trials where the object to be guessed was a number of two digits—*i.e.* one of the 90 numbers included in the series 10 to 99—chosen at random by the agent. Out of 2,624 trials, where the most probable number of successes was 29, the actual number obtained was no less than 275.” After observing that there were a good many more correct reverse-order guesses, we read that “the argument for thought-transference afforded by the total of 275 cannot be expressed here in figures, as it requires 167 nines—that is, the probability is far more than the ninth power of a trillion to 1” (a European trillion, rather than an American one).

In Volume II we have the “cases too late for insertion in their proper places,” where we learn more about the Misses Wingfield: “Miss M. Wingfield, sitting six feet behind the percipient, drew a slip at random, and fixed her attention on the number which it bore; Miss K. Wingfield made a guess at the number, and the real number and the guess were at once recorded in the Table. The slip of paper was then replaced, the contents of the bowl shuffled, and another draw made at hap-hazard.” Miss K. Wingfield made 27 hits in 400 trials, and in 21 cases gave the digits in reverse order; in 162 other cases one of the digits was in its right place. The writer concludes that the odds here “are nearly two hundred thousand million trillions of trillions to 1.”

To these odds we shall return, but first a word on Kate Wingfield. She was a

<sup>43</sup> “Messrs Hansen and Lehmann on the Telepathic Problem,” *J. Soc. Psych. Res.*, 1899, 9:113–120, on p. 120. The short essays and correspondence of most of the protagonists in this debate, together with references to the rest, may be conveniently read in James, *Essays in Psychical Research* (cit. n. 21), Ch. 22, “Controversy with Titchener,” pp. 167–179, 480–484.

<sup>44</sup> Edmund Gurney, Frederick W. H. Myers, and Frank Podmore, *Phantasms of the Living*, 2 vols. (London: Trübner, 1886).



protégée of Myers, who encountered her in 1884. We have not the slightest reason to doubt that a process of randomization was used in this case, for the technique had been promulgated by Richet, and Myers would assuredly have commended this method to the Wingfield girls. On the other hand, we now well know the difficulty of randomizing slips of paper that are extracted and replaced. Bowls filled with pieces of paper have “number habits.” Myers provided no information on the number of occasions on which the Wingfields conducted such experiments. If they were performed on a number of occasions, and if, moreover, Miss K. Wingfield heard Miss M. Wingfield indicate on each trial if she were successful—so that a good run would be continued and a bad one discontinued—the success of Miss K. Wingfield on two experiments is less astonishing by far than Myers imagined. There is no need to postulate any conscious fraud whatsoever, and we may suppose that these incidents of randomization teach us more about defective randomizers than anything else.

It should also be recorded that Kate Wingfield appears to drop out of the psychic world, but did not in fact do so. She is the “Miss A.” who occurs repeatedly in communications to the SPR, chiefly from Lady Radnor. She was a welcome member of séances at country houses and gradually assumed a complete repertoire, passing from rapping, table lifting, and automatic writing to crystal balls and, in the end, to serving as a full-trance medium.<sup>45</sup> Some forty years after her first experiments in number guessing she released to the world specimens of her automatic writing, *Guidance from Beyond*.<sup>46</sup>

#### VIII. PEIRCE'S CRITICISMS OF RANDOMIZATION IN PHANTASMS

An American Society for Psychical Research was established in 1884, just two years after the London society, and just as Peirce and Richet were publishing their papers inaugurating the practice of randomization. Its first project was a census of telepathic experiences. As in England, physical scientists figured prominently. Until its twentieth-century reorganization, the American society published only one volume, for 1885–1889. It bears upon its title page the possibly telling misprint *Proceedings of the American Society for Physical Research!*

Membership was heavily concentrated in Boston for reasons that, to the modern reader, are most easily gleaned from Henry James's novel *The Bostonians*. Boyleston Street, Marlborough Street, Commonwealth Avenue, and the like figure heavily among the addresses of members and associates. The first president was Simon Newcomb, and the vice presidents included another astronomer (E. C. Pickering), two Harvard medical men (H. P. Bowditch and Charles Sedgwick Minot), and two philosopher-psychologists (G. Stanley Hall and G. S. Ful-

<sup>45</sup> F. W. H. Myers wrote of Wingfield in the *Proceedings* under titles such as “The Subliminal Consciousness,” 1892, 8:436–535, and 1893, 9:26–128, but he called her “Miss A.” Cf. Myers, *Human Personality and Its Survival after Bodily Death*, 2 vols. (London: Longmans, Green, 1903). See also Sir Lawrence J. Jones, “Presidential Address,” *Proc. Soc. Psych. Res.*, 1928, 38:17–48, the topic of which is K. Wingfield; and Allan Gauld, *The Founders of Psychical Research* (London: Routledge & Kegan Paul, 1968), pp. 223f., 246–248.

<sup>46</sup> *Guidance from Beyond, Given Through K. Wingfield with a Preface by Helen, Countess of Radnor, and Introduction by Sir E. Marshall Hall* (London: P. Allan, 1923); and *More Guidance from Beyond* (London: P. Allan, 1925). See also H. M. R. [Helen Matilda, Countess of Radnor], ed., *Teachings: Automatic Writings by Catherine Elizabeth Wingfield* (London: Women's Printing Society, ca. 1928).

lerton, the latter one of the authors of the essay on "small differences" cited above). William James and Josiah Royce were among the more active participants. Despite the enthusiasms of James, the Boston society was a great deal more skeptical than its London counterpart. By 1889 it had concluded (with Richet) that there was no sound statistical support for telepathy. The society became insolvent and in effect disbanded in order to become a branch of the SPR in London. The branch secretary was the generous enthusiast Richard Hodgson, who gave it his all not only up to his death on 20 December 1905 but also thereafter, by engendering William James's most extensive investigations of the paranormal.

The most acrimonious exchange in the old American *Proceedings* was between Peirce and Gurney. Peirce was neither friend nor member of the society, but he challenged *Phantasms of the Living* as being a gross misapplication of probability logic. I have cited a couple of the probability claims of *Phantasms* in connection with the Misses Wingfield. There are a great many more that provoked Peirce's tirade. The trio of authors "cipher out some very enormous odds in favor of the hypothesis of ghosts. I shall not cite these numbers, which captivate the ignorant, but which repel thinking men, who know that no human certitude reaches such figures of trillions, or even billions, to one." Peirce then proceeds to state the logical form of Gurney's argument in probabilistic thinking and to argue that every single cited coincidence "sins against one or more of sixteen different conditions to which such an argument must conform to be valid." The debate with Gurney occupies sixty-five pages of criticisms and rejoinders.<sup>47</sup> It is Peirce at his crankiest (but none the less sound for that). One suspects that many of the Boston skeptics were egging him on.

#### IX. STANFORD'S COOVER

Richet initiated the use of physical randomizers in testing the hypothesis of widespread low-level telepathic powers. Like the American investigators, he concluded within five years that no evidence for such powers can be discerned by these means. Although he began the use of randomization in empirical experimentation (experimentation where we have little idea of what we are investigating) independently of, if at the same time as, Peirce, he did not develop what I have called randomized experimental design. Peirce did, but his innovation was ignored by psychologists. Hence a further event in parapsychology usefully rounds out my story. By chance it began in 1884, the year of Peirce's and Richet's publications.

That was the year in which, halfway round the world, Leland Stanford, Jr., died of typhoid fever at the age of seventeen. His parents, deciding to make all the children of California their children, established Leland Stanford Junior University in 1885. As befitted her era and her station in life, Mrs. Leland Stanford was a spiritist and spiritualist who longed for and believed in communication from the dead, especially Leland, Jr. So too was Leland Stanford's brother,

<sup>47</sup> C. S. Peirce, "Criticism on *Phantasms of the Living*: An Examination of an Argument of Messrs. Gurney, Myers, and Podmore," *Proc. Amer. Soc. Psych. Res.*, 1885-1889, 1:150-157, on p. 150; Edmund Gurney, "Remarks on Professor Peirce's Paper," *ibid.*, pp. 157-179; and Peirce, "Mr. Peirce's Rejoinder," *ibid.*, pp. 180-215.

Thomas Welton Stanford, who had migrated farther west, to Melbourne, Australia. There he had a vigorous circle of spiritists. Their primary interest appears to have been “aports,” material objects that were transported into a séance room by spirits. He donated £10,000 to the young university as an endowment for psychic research.

The university was not quite sure how to use the money, but in 1912 it was placed at the disposal of the psychology department, then rapidly establishing itself as a leading center for the subject. There had been some unfortunate spiritualist donations to other American universities, most notably the University of Pennsylvania, and the Stanford department hesitantly sought advice from such earlier beneficiaries. The head of the department noted an additional problem: “The situation was further complicated in the country at large and especially in California by the presence among the devout Spiritualists of many false teachers who sought to exploit spiritualistic procedure for pecuniary profit.” T. W. Stanford’s gift was unconditioned, and the psychology department found itself able to establish a Division of Psychical Research with a well-equipped laboratory. A fellow in psychic research was appointed and made director of this laboratory: “To dignify the fellowship in the regard of the university world, the Trustees conferred on [this appointee] Dr. Coover the rank of Assistant Professor.”<sup>48</sup> In fact John Edgar Coover (1872–1938) had been an assistant on an annual appointment in the psychology laboratory since 1910. He remained the fellow until 1937 and rose through the ranks to full professor. There is no evidence that he did a stroke of psychical research after 1917, and thus was T. W. Stanford’s fund effectively nullified and absorbed into general revenue.

That endowment did, however, lead to intensive work between 1912 and 1917, when it culminated in *Experiments in Psychical Research at Leland Stanford Junior University*. This 641-page book, the first (and last) of the proposed series of psychical research monographs, is a remarkable application to its subject of the evolving techniques of American university psychology of the day. Hordes of undergraduates in the large psychology classes were made into the subjects for psychological testing.

I shall describe only two bodies of experiments on thought transference or telepathy, but there is a vast range covering items from “the feeling of being stared at” to subliminal perception. In the latter case there is a remarkable study of sound assimilation, addressing the question: “*Given the English language spoken under conditions adequate for communication, to what extent does perception of words depend upon the sense of hearing, to what extent upon what the mind supplies?*”<sup>49</sup> Over four years, 40,500 experiments using 107 students were conducted with nonsense syllables. One point was to determine the extent to which people heard nonsense syllables as significant, and hence to suggest inferences about the ways in which members of a séance might “hear” the mumblings of a medium through a trumpet as naming dead friends or family whom the medium could not possibly have known about.

Telepathy was investigated in all the standard ways, with particular attention

<sup>48</sup> F. Angell, “Introduction,” in John Edgar Coover, *Experiments in Psychical Research at Leland Stanford Junior University* (Psychical Research Monographs, 1) (Stanford, Calif.: Leland Stanford Junior University Publications, 1917), pp. xix, xxii.

<sup>49</sup> Coover, *Experiments*, p. 371 (italics in original).

to guessing playing cards. Here there were two groups of experiments, one using students and one using amateur or professional psychics. In the former case there were 97 agents and 105 reagents. The agent, it will be recalled, picks a card at random and in various ways thinks about it, while the reagent states what card it is. These 202 investigators were from the general lecture course in psychology, augmented by a few, such as Coover, who were engaged in laboratory work in the psychology department. They were majors in twenty different subjects at Stanford and represented twenty-one states in addition to the several regions of California. All were chosen on the condition that they were generally favorable to the hypothesis of telepathy and similar topics of spiritist research. Only four were willing agnostics who were open to the possibility of psychic phenomena. After four or five days of experimentation the reagents in general were not discouraged by the boring routine and expected that they had scored quite well. That was how it *felt*.

Ten thousand trials were made on these subjects. Conditions were varied in numerous ways. The agent was instructed to “image” a drawn card in one of three different ways: visually, kinaesthetically (i.e., by silently uttering the description of a drawn card), and by both together while hearing the name of the card in the imagination. The kinaesthetic mode was an attempt to test Hansen’s 1895 hypothesis of involuntary whispering.

In the first 3,000 experiments, the agent changed the distance from the reagent every 20 trials, in order to see if it helped to be close or far away. In the next 2,500 trials the duration of imaging varied between 20 and 60 seconds, with the reagent being allowed to choose the time for some of the series.

For our purposes, however, the most notable innovation was the use of randomized control against regular experiments. In the regular experiment the agent drew a card at random and imaged it. In the control experiment the reagent guessed, but the agent did not look at the card at all. It was decided at random whether the trial would be regular or control. In Coover’s description:

The experimenter with a watch before him, (1) shuffles the deck of 40 playing cards (the face cards being discarded),<sup>131</sup> cuts the pack, and holds cards concealed; (2) shakes the dice-box, to determine a control or regular experiment, and, if the latter, the form of content the card is to have in his mind; (3) if a regular experiment, he turns over the pack, exposing to his view the under card, taps once to signal the reagent that the experimental period begins, holds mental content of card and wills the content to be projected into the mind of the reagent, and after 15 to 20 or more seconds<sup>132</sup> taps twice to signal the close of the interval. After he notes that the reagent has recorded his guess, and has turned to his introspections, he records the color, number, and suit of the card and the number of the die-spot which conditioned the form of the experiment (as, *R5H 1*, for Red, Five of Hearts, Die-spot 1—i.e., held in Visual Impression). The control experiments ran off in precisely the same form as the regular, except that the card remained unknown until the reagent had recorded his guess.<sup>50</sup>

The dice box determines the type of “treatment.” Outcomes 1, 3 and 5 denote the choice of method of imaging the card (visual, kinaesthetic, or everything).

<sup>50</sup> *Ibid.*, p. 53f. Footnotes 131 and 132 in the text refer to the American SPR’s instructions of 1886 for the conduct of the telepathy census. The society had proposed matched experiments, in which occasions when the agent knew the card guessed were matched with occasions when the agent knew nothing and so there was no knowledge to transmit to the reagent.

Outcomes 2, 4, and 6 indicate that a control experiment is to be done. Thus in Coover's design we have both the use of the randomizer (playing cards) for choosing which object is to be guessed and the use of the randomizer (the dice box) to assign one of three possible types of treatment or a control experiment in which there is no treatment.

Here we have the whole retinue of control, blindness, and randomization. It is not for the first time, even in Coover's own researches. His very first experiment in this series, conducted in 1912, prefigures the work with cards, although it involved only one experimenter, Coover, and one reagent, an advanced psychology student. The aim was identical with that of the experiment of the Wingfield sisters, so much admired by Myers but about which I have expressed reservations. The reagent aimed at guessing couplets of numbers from 10 to 99. But instead of using slips of paper, as the Wingfields had, the numbers were chosen at random from "lotto blocks." Lotto was a predecessor of bingo by several centuries and was in the United States more commonly called keno after the 1870s. It is, under the name "lotto," widely in use in state lotteries, while as "keno" it was a recent favorite of competitions for newspaper promotions. Traditionally balls with the numerals from 10 to 99 were used, and a player had to guess five or seven that were drawn at random. Although no amount of vigorous shaking and shuffling of an urn full of ninety balls or blocks will guarantee a great irregularity or freedom from pattern in a long sequence of draws with replacement, it is a far better guarantee than the drawing of ninety slips of paper from a bowl.

Our interest is less in the blocks as an improvement over slips of paper than in their use in combination with a dice box. Everything is now randomized: both the choice of number as "treated" (i.e., known and thought about) or control and the choice of method of transmission. Randomization was coming of age in American departments of psychology, although nowhere, I believe, was it more systematically practiced at such an early date than in this subdepartment of psychological research.

The experimenter sat facing the reagent's back, drew from a bag a Lotto-Block, and, if the numbered side came up, made ready to hold it in some kind of vivid imagery, shook the dice-box, tapped with his pencil once to inform the reagent that the period of impression began, then held or did not hold imagery of the number; and after 15 seconds tapped twice to close the period of impression. When the numbered side of the block came up, imagery was held according to the face cast by the die, as follows: (1) Visual impression; (2) Kinaesthetic image (care being taken to avoid an auditory accompaniment, and also any movements of pronunciation great enough to be felt); (3) Auditory imagery (stripped of its usual kinaesthetic accompaniment); (4) Combination of 1 and 2; (5) of 1 and 3; (6) of 2 and 3. In order that there would be no confusion, the experimenter kept before him a card with the kinds of impressions, or imagery, tabulated by number, upon it. In case the blank side of the block was drawn, the dice-box was shaken, and the experiment progressed in every way like its alternate; except that the experimenter refrained from thinking of numbers (by musing upon an ocean scene).<sup>51</sup>

It may be instructive that in his description of this inaugural experiment, conducted in 1912–1913, Coover uses the terminology of regular and control experi-

<sup>51</sup> *Ibid.*, p. 36.



ments; he puts the word *control* in what philosophers more recently have called “scare quotes,” that is, he writes of a “‘control’ experiment.” This may indicate a warning that he is using a novel form of expression. In reporting subsequent experiments he drops the scare quotes. Much earlier, people had talked of having adequate control for experiments—even in the *Proceedings* of the American SPR and the 1835 referee’s report quoted in note 10 above. It would be interesting to speculate, however, that it is Coover to whom we owe the explicit terminology of controlled and control experiments.

Coover’s conclusions were clear: “That no trace of an objective thought transference is found either as a capacity shared in a low degree by our normal reagents in general (Richet’s ‘Suggestion Mentale’) or as a capacity enjoyed in any perceptible measure by any of the individual reagents.”<sup>52</sup> It was also noted that for reagents who had an inner experience of being right on certain occasions there was no correlation between their flashes of certainty and the correctness of their guesses.

Coover’s statistical analysis was vastly better than Richet’s but still drew largely on Edgeworth’s techniques, to which I have alluded. I should note one problem, that of “scoring.” The reagents guessed several aspects of a card: its suit, its color, its number. Naturally it is of interest whether colors are transferred better than numbers, or suits better than colors, for in this purely empirical investigation we have no idea what is being transferred. A partially correct guess might be enormously interesting (suppose an agent got the colors right very often, although being sadly inadequate for suits and numbers). How does one score partial success, so as to bring out such questions of interest? Coover had a mere commonsense approach to what is a nice problem in combinatorics. The mathematical solution was provided by R. A. Fisher as a favor to workers in the SPR.<sup>53</sup>

#### X. FISHER, THE SPR, AND TASTING TEA

Thus we return to R. A. Fisher, the acknowledged founder of randomized experimental design. It is not my contention that he learned from Coover in any interesting way. He knew of this kind of work, but a very different level of sophistication was involved in Fisherian randomization. As I said at the outset, Fisher’s claims for randomization were disputed and are disputed. They are located within his own vision of the conceptual foundations of statistics. He held, first, that the introduction of physical randomizers into an experimental design increased or perhaps even made possible an objective measure of the outcome of the experiment. Second, randomization increased the efficiency of an experiment in the sense that fewer trials were necessary in order to achieve a given level of significance. In the case of traditional statistical analysis in agriculture, using balanced designs and error analysis, one got less probable error (so Fisher argued) with fewer plots and hence less cost if one randomized. Such ideas have roots, insofar as they are not Fisher’s own inventions, in bits of the history of data analysis that have nothing to do with Coover.

Nevertheless, Fisher was well aware of psychical research, even if, on the

<sup>52</sup> *Ibid.*, p. 124.

<sup>53</sup> R. A. Fisher, “A Method of Scoring Coincidence in Tests with Playing Cards,” *Proc. Soc. Psych. Res.*, 1924, 34:181–185.

testimony of the daughter who is his biographer, he thought it baloney.<sup>54</sup> I have noted his 1924 paper scoring partial success in guessing cards. There is a further contribution in 1928, an appendix to a paper by Ina Jephson. Here the problem is once again one of scoring. How does one allow for the "number habits" first noticed by Minot in 1886? Jephson in effect repeated and improved on Coover's experimental work. "Special thanks," she writes, "are due to R. A. Fisher for devising the scoring system which was used throughout this experiment, and which made it possible to arrange such a large number of different elements on a business-like basis, and also for having many times flung me a life-line when I was nearly submerged by the problems which taxed my limited arithmetic." Fisher characteristically did not merely solve the mathematical problem, but provided a technique by which the fieldworker could do the computations. When the results were criticized, Fisher, once again characteristically, wrote in spirited defense of the person whom he had aided.<sup>55</sup>

These little papers were written about the time that Fisher was forming his own ideas on the importance of randomization and beginning to campaign for its use in experimental agriculture and elsewhere. I do not assert that the forty-year history of randomization in parapsychology was an important "influence" upon Fisher. To do so would be to misunderstand the core of Fisher's rationale in terms of maximizing information and obtaining objective measures of reliability. I do note that Fisher was intimately aware of the practices and the past of psychical research. I allow myself to conclude with one whimsical speculation. It concerns the heroine of *The Design of Experiments*.

Every student of experimental design or of randomization is acquainted with the figure in the first substantial chapter of that book, the lady tasting tea. She asserts that she can tell the difference between a cup of tea to which milk has been added after pouring and a cup of tea with the milk in first. Fisher proposes that she be given eight cups to taste and recommends that the treatment of each cup be decided at random. There are two possibilities. One is that a total of four milk-in-first cups be used, presented with the other four in random order. Fisher prefers the other possibility: that the mixture in each cup be decided at random, so that there might well be five or six milk-in-first cups, and only two or three of the other sort. He grants, however, that the lady might claim that a longish chance run of this sort might impair her ability to make the necessary discernments. It will be recalled that this worry was expressed in the very first randomized experiment, that of Peirce and Jastrow.

We do very probably know the identity of the lady who was subject of the tea-tasting story: Muriel Bishop, a student of algae at Rothamstead.<sup>56</sup> She was offended at being offered a cup of tea into which milk had been poured before the tea. Could she really tell the difference? And so the test was conducted, although we do not know the results.

<sup>54</sup> Box, R. A. Fisher (cit. n. 4), p. 237.

<sup>55</sup> I. Jephson, "Evidence for Clairvoyance in Card Guessing: A Report on Some Recent Experiments," *Proc. Soc. Psych. Res.*, 1928, 38:223–269, on p. 223. R. A. Fisher, "The Effect of Psychological Card Preferences," *ibid.*, pp. 269–271; and, for the defense of Jephson, Fisher, "The Statistical Method in Psychical Research," *ibid.*, 1929, 39:185–189. For a short survey of the good and bad things that statisticians have done for parapsychology see Persi Diaconis, "Statistical Problems in ESP Research," *Science*, 1978, 201:131–136, on p. 133.

<sup>56</sup> Box, R. A. Fisher (cit. n. 4), p. 134.

The ability to discern the order in which milk is added to tea may seem curious to those ignorant of the fine cues to class distinction in England. The upper classes pour first the tea and then the milk, while the lower classes do it in the reverse order. In the 1950s, when Nancy Mitford made popular many of these distinctions under the rubric “U” or “non-U,” the term *miffy* came into vogue to denote those who put their milk in first.<sup>57</sup> We may also note that, formally speaking, the problem of distinguishing mif from non-mif tea is identical with that of the reagent asked to distinguish red from black cards. In his chapter Fisher proceeds with several little jokes, for example, testing a person for more possibilities, India or China tea, which may be either mif or non-mif. Can one not, without excessive whimsy, hear the distinction China plus mif (for example) as black plus face card, with the ensuing problem of scoring?

At any rate the social setting is right. It is not just that the tea taster and the sensitive alike are typically female. The sensitives of Fisher’s day and before were genteel women, typically associating a little above their station and hence attentive to the delicacies of life. One has some confidence that the Misses Wingfield, for example, would never have put their milk in first. Anyone unfamiliar with the social combination of séance and tea may consult the elegant photograph *After the Séance*.<sup>58</sup> Fisher was not unknown to play quiet and complex jokes upon the learned, and upon those whom he spoofed, simultaneously.

<sup>57</sup> Nancy Mitford, *Noblesse Oblige: An Enquiry into Identifiable Characteristics of the English Aristocracy* (New York: Harper, 1956).

<sup>58</sup> See Ronald Pearsall, *The Table Rappers* (London: Michael Joseph, 1972), facing p. 209.